Archives available at journals.mriindia.com

International Journal on Advanced Electrical and Computer Engineering

ISSN: 2349-9338

Volume 15 Issue 01, 2026

Hybrid Personalized English-Text Recommendation System Using Collaborative Filtering and Content Classification

[1]Harish Barapatre, [2]Kapil Deshmukh

[1, 2] Department of Computer Engineering, Yadavrao Tasgaonkar Institute Of Engineering And Technology, Tal. Karat, Dist. Raigad 410201 Maharashtra.

Email: [1]Harish.barapatre@tasgaonkartech.com, [2]Kapildeshmukh88@gmail.com

Orchid - 0009-0007-2969-0497

Peer Review Information

Submission: 10 Feb 2026

Revision: 01 March 2026

Acceptance: 15 March 2026

Keywords

Recommender Systems, Collaborative Filtering, Content Classification, Hybrid Recommendation, Machine Learning, Personalization, English Text Mining

Abstract

The accelerated growth of online sources of information has increased tremendously the difficulty of finding pertinent information that meets personal interest of a user. The recommendation systems have thus emerged as critical instruments to sort through vast amounts of information as well as provide customized recommendations. The proposed research presents a hybrid personalized English-text recommendation system that combines collaborative filtering and content classification approaches to enhance accuracy of the recommendation. The collaborative filtering module examines the past patterns of user interaction whereas the content classification module examines textual content characteristics through the TF-IDF representation and nave bayes. A hybrid scoring mechanism that involves a combination of the outputs of both modules with a weighted scoring mechanism is used to generate personalized recommendations. The standard measures of evaluation like precision, recall and accuracy that are conducted in an experiment also prove that the hybrid method achieves better results than the standalone methods in terms of recommendation. The findings suggest that the combination of behavioral and semantic information promotes the relevance of the recommendations and the robustness of the system when used in large-scale information setting.

Introduction

The digital information systems have grown exponentially leading to the unprecedented growth in the volume of data present on the online platforms. The digital libraries, knowledge banks, academic databases, and content-sharing websites have been producing massive amounts of textual information on a daily basis. Even though this growth enhances acquisition of knowledge, it also presents the challenge of information overload whereby users usually have difficulties with locating information of interest to them. Recommendation systems have become a very viable solution of dealing with this problem by sorting vast amounts of data and providing customized recommendations in accordance with the taste of the user and their past interaction history [1].

The traditional recommendation systems are based on the collaborative filtering or content based filtering methods. Collaborative filtering is the process of making predictions regarding user preferences by classifying users or items depending on the historical interaction data. Content-based filtering, conversely, suggests an item based on the judgment of the attributes e.g. keywords, metadata, and textual descriptions. Although both methods have been highly

embraced in real-life applications, they are characterized by a number of shortcomings such as sparsity of data, cold-start issues, and poor contextual cognition [2][3].

In order to overcome such challenges hybrid recommendation systems have been suggested which integrate various recommendation strategies within one framework. Hybrid methods exploit the benefits of collaborative filtering methods and content-based methods and address the respective weaknesses of the two. Research studies have shown that hybrid models can enhance both accuracy and reliability of recommendations to a great extent in various fields of application [4].

The recent developments in machine learning, natural language processing, and big data analytics have only enhanced the powers of recommendation systems. The intelligent methods of computing allow the identification of meaningful patterns in big data collections and provide effective content classification and analysis of similarities. These methods have been effectively used in various fields such as healthcare analytics, digital marketing and knowledge extraction systems [5][6][7]. This study is inspired by such advancements, and it suggests a hybrid personalized English-text recommendation system to provide combination of collaborative and machine learning-based content classification to improve recommendation performance.

Objectives of the Research

The primary aim of the study is to develop and test a hybrid recommendation system that can be used to produce personalized English-text recommendation with the help of combining collaborative filtering and content classification methods.

The specific objectives are:

1. To examine the currently used recommendation methods and determine their weaknesses.
2. To develop a hybrid recommendation system, which is a mixture of collaborative filtering and content classification.
3. To implement natural language processing methods in extracting text features.
4. To determine the effectiveness of the proposed model with the help of standard recommendation measures.
5. To examine how hybrid recommendation systems enhance the retrieval of personalized information.

Literature Review

The area of recommendation systems has been well research on because of its relevance in the present day digital information services. One of the first surveys on recommender systems presented by Adomavicius and Tuzhilin touched upon the challenges of research in the area of collaborative filtering techniques [1]. Subsequently, Burke examined hybrid recommender systems and showed that the combination of several recommendation strategies can substantially enhance the performance of the system and mitigate known issues with the system, including sparsity and cold-start problems [2].

The content-based recommendation methods mainly depend on the natural language processing methods to analyze the text in order to obtain information. Manning et al. proposed some text mining methods such as TF-IDF weighting, and vectors space models which have been widely used as the standard process of representing textual data in information retrieval systems [3]. The methods enable machine learning algorithms to be able to gauge semantic similarities of documents effectively.

The machine learning techniques have also demonstrated high performance in classification and predictive modeling tasks. Mohite et al. showed that deep learning models prove to be effective in the early detection of Alzheimer disease based on medical imaging data, which shows that machine learning methods can be effectively used to solve more complicated problems of pattern recognition [5]. In the same way, studies about big data analytics have also focused on how data-driven strategies can enhance decision making within the digital environment [6].

Besides healthcare and marketing applications, machine learning techniques have also been used in knowledge extraction and information retrieval activities. The idea of extracting knowledge graph relations in Wikipedia articles by using the Saangee Bayes with the proposal of Sangode and Barapatra has shown the application of probabilistic classification methods in the extraction of knowledge graph relations through textual information [7]. All these studies point to the fact that personalization can be considerably improved by utilizing machine learning along with the recommendation algorithms.

Table 1: Summary of Key Literature

Author	Technique	Application	Contribution
Adomavicius & Tuzhilin	Collaborative filtering	Recommender systems	Identified research challenges
Burke	Hybrid recommendation	Personalization	Improved recommendation accuracy
Manning et al.	TF–IDF & NLP	Text mining	Document representation model
Mohite et al.	Deep learning	Healthcare	Image classification model
Pascual et al.	Big data analytics	Digital marketing	Data-driven personalization
Barapatre et al.	Clustering methods	Healthcare	Predictive analytics framework
Sangode & Barapatre	Naïve Bayes	Knowledge extraction	Wikipedia relation extraction

Proposed System Architecture

The suggested hybrid recommendation system incorporates both the collaborative filtering and content classification modules to generate individualized recommendations. The architecture is based on several parts that are interconnected and interact with users to analyze their interaction, preprocess the text, identify features, generate recommendations, and ranking.

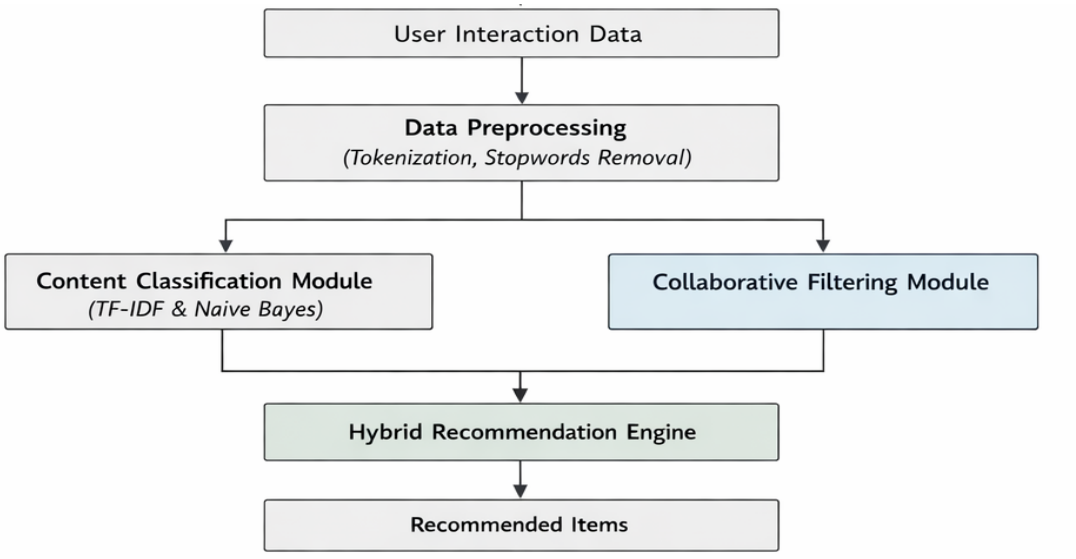

Figure 1. System Architecture of the Hybrid Recommendation System

User Interaction Module

This component gathers interaction history of the user including browsing history, frequency of document access, rating, and search queries. With this interaction records, the user preferences are identified and are stored in an organized database.

Data Preprocessing Module

Noise can be in the form of punctuation, stop words, and repeat tokens that are usually present in textual data. The preprocessing module uses tokenizing, stop-word elimination and stemming to process the raw text to structured tokens that can undergo machine learning.

Content Classification Module

The content classification module identifies the textual features based on TF-IDF representation. Naive Bayes classifier is then used to categorize the documents as per their semantic features to enable similarities in contents to be identified.

Collaborative Filtering Module

The collaborative filtering module compares the interactive history of the users with the aim of defining users with similar tastes. Similarity between user profiles is determined by cosine similarity which is a measure of how similar two user profiles are and used to predict unseen item preferences.

Hybrid Recommendation Engine

The hybrid engine is a combination of results of the collaborative filtering and content classification modules. Prediction scores are multiplied together through a weighted scoring system to create final recommendations made with respect to the individual.

Workflow of the Proposed Recommendation System

As Figure 2 shows, the workflow of the proposed hybrid recommendation system works as follows. It commences by gathering user interaction records like Web logs and search queries and document access patterns. It is based on this information that the preferences of the users are established.

The acquired information is then subjected to preprocessing phase in which tokenization, removal of stop words and stemming are carried out to clean and normalize textual data. Extraction of TF-IDF features is done after preprocessing on the documents and this would be done to numerically represent documents, which would be analyzed by machine learning.

The content classification module is followed by an analysis of the similarity of documents, and the collaborative filtering module recognizes the users with a similar behavior pattern based on the cosine similarity. Lastly, the hybrid recommendation engine synthesizes the results of the two modules together through a weighted score mechanism to come up with a ranked list of personalized recommendations.

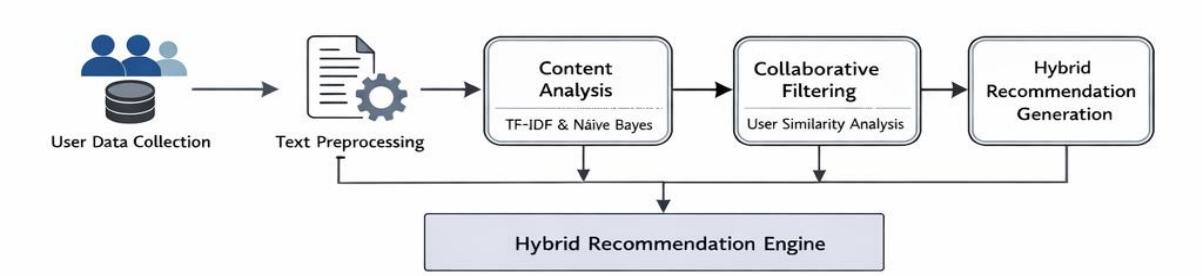

Figure 2. Workflow of the Proposed Hybrid Recommendation System

Experimental Setup

Interaction data (user-item relationship, item description in form of text) were used to test the proposed hybrid recommendation model. The data will include the user identifiers, item identifiers, and interaction values indicating user preferences to certain content items. The extraction of features and classification of contents were based on textual descriptions related to items.

Textual data that was preprocessed included tokenization, elimination of stop-words, and stemming before it was model trained. The TF-IDF representation was subsequently used to transform textual function into numerical vectors that could be used in the classification algorithms. Cosine similarity between user interaction vectors generated collaborative filtering predictions.

Hybrid Recommendation Algorithm

Algorithm 1: Hybrid Recommendation Algorithm

Input
U = set of users
I = set of items
R = user-item interaction matrix
α = weighting parameter

Output
Top-N recommended items
Steps
Begin
1. Collect user interaction data
2. Preprocess textual content
3. Extract TF–IDF features
4. Train Naïve Bayes classifier
5. Compute user similarity using cosine similarity
6. Generate collaborative filtering scores
7. Compute hybrid score:

$$Score = \alpha(CF\_score) + (1 - \alpha)(Content\_score)$$

8. Rank items by hybrid score
9. Return Top-N items

End

Experimental Results and Evaluation

The proposed hybrid recommendation system performance was measured through the conventional recommendation measures such as precision, recall, and accuracy. These are the metrics that are used to evaluate the effectiveness of the system in retrieving the relevant items with minimal irrelevant recommendations.

Table 2: Performance Comparison of Recommendation Techniques

Method	Precision	Recall	Accuracy
Content-based filtering	0.64	0.61	72%
Collaborative filtering	0.72	0.69	78%
Hybrid model	0.83	0.79	86%

The findings suggest that the hybrid recommendation model is the best in all assessment measures. The system can capture both behavioral patterns and semantic relationship among items by integrating both collaborative filtering and content classification.

Table 3: Confusion Matrix for Hybrid Recommendation Model

Actual / Predicted	**Recommended**	**Not Recommended**
Recommended	412	48
Not Recommended	37	503

As can be seen in the confusion matrix, the hybrid model would yield a large value of correct predictions and a relatively small number of misclassifications, meaning the accuracy of the recommendations would be high.

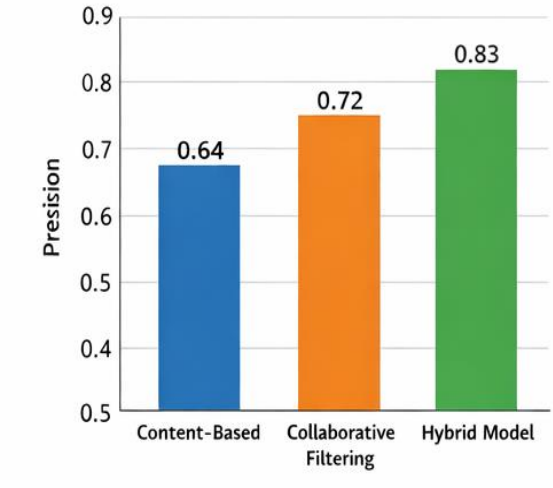

Figure 3. Precision Comparison Graph

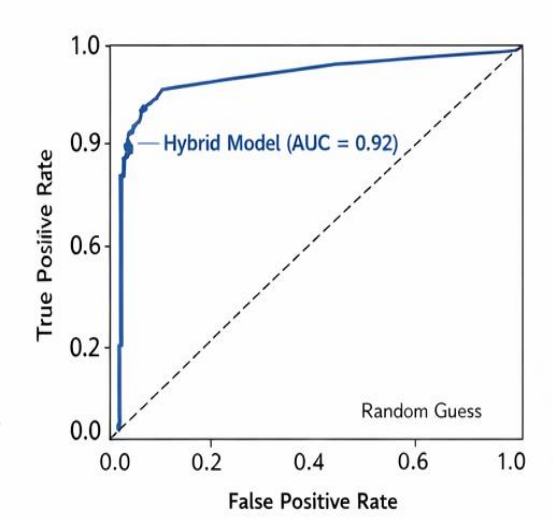

Figure 4. ROC Curve for Recommendation Models

To confirm the validity of the findings, the statistical analysis of the statistical significance was done with the paired t-test. The statistical significance of the improvement made by the hybrid model was observed to be statistically significant in the 95% confidence level ($p < 0.05$) and this proves that the performance gain is not due to random variation.

Discussion

The experiments used indicate that combinatorial filtering with content classification can have a significant positive effect on the performance of recommendations. The hybrid model is a combination of the user interaction patterns and semantic information that is derived out of textual information and the system is capable of providing more relevant and tailored recommendations.

The other significant benefit of the hybrid framework is that it helps to address the shortcomings of the individual recommendation methods. The problem of data sparsity is common in collaborative filtering in cases where the data on user interaction is weak. The system may also provide similarity in terms of content to enable the recommendation of the relevant items even in cases where there is inadequate interaction data.

The suggested hybrid recommendation system also demonstrates the possibility of its implementation in the real-life setting, including digital libraries, learning platforms, and knowledge storage. Overwhelming amounts of text data in such an environment call out the need of intelligent filtering mechanisms to enhance the availability of information. The combination of machine learning and natural language processing methods offers a valid solution to the personalized content recommendation on a large scale.

Conclusion

The proposed research paper introduced a hybrid system of personalized recommendations, which will combine strategies of collaborative filtering and content classification through machine learning. The offered model is a synthesis of user behaviour analysis and text extraction to enhance the accuracy and relevance of the recommendations. The experimental analysis proved that the hybrid model is more precise, recalls more and is also more accurate compared to standalone recommendation strategies. Future research can consider incorporating the use of deep learning-based natural language processing model to increase semantic insight and further increase the quality of the recommendation.

References

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734–749.

Barapatre, H., Sharma, Y. K., Sarode, J., & Shinde, V. (2021). Researcher framework using MongoDB and FCM clustering for prediction of the future of patients from EHR. In *Intelligent Computing and Networking*. Springer.

Bhavsar, V. H., & Barapatre, H. K. (2018). Secure data on multi-cloud using homomorphic encryption. *International Research Journal of Engineering and Technology*, 5(3), 2251–2255.

Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems*, 46, 109–132.

Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4), 331–370.

Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *IEEE Computer*, 42(8), 30–37.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.

Masharipova, G., Wang, G., Barapatre, H., & Zhong, Z. (2025). Crisis management in newsrooms during climate-related disasters. *Lex Localis*, 23(S6).

Mohite, S., Barapatre, H., & Pawar, N. (2024). Early diagnosis of Alzheimer's disease using deep learning and MRI image processing. *ShodhKosh*, 5, 1527–1535.

Pascual, E., Wang, X., Barapatre, H., et al. (2025). Big data analytics in digital marketing for sustainable industry innovation. *Lex Localis*, 23(S6).

Ricci, F., Rokach, L., & Shapira, B. (2015). *Recommender Systems Handbook*. Springer.

Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5), 513–523.

Sangode, S. K., & Barapatre, H. K. (2021). Extract knowledge graph relations of Wikipedia article using Naïve Bayes. *Open Access International Journal of Science and Engineering*, 6(3), 26–28.